# APPENDIX F-1

*Estimating the Benefits of Abating Toxic Air Pollutants: What do Benefits Assessors Need from Risk Analysis – and Why are they Unlikely to Get it?*

White Paper by Dr. Lester Lave,
Graduate School of Industrial Administration,
Carnegie-Mellon University, Pittsburgh, PA.

# Estimating the Benefits of Abating Toxic Air Pollutants: What Do Benefits Assessors Need From Risk Analysis – and Why are They Unlikely to Get It?

Lester B. Lave 5-22-00
Carnegie Mellon University

## Abstract

I examine what benefits assessors require to estimates of benefits and costs of abating hazardous air pollutants, why current toxicological information is not suited to benefits estimation, and a research agenda that would lead to confident estimating of the benefits of abatement. While it has many limitations and uncertainties, epidemiology data provides information for benefits assessment. In contrast, regulatory toxicology is focused on protecting people from harm, not on giving best (unbiased) estimates of the harm to people from various exposures. Few toxicology tests have been validated in the sense showing that they predict human toxicity. Other than reasoning from first principles, we have no way of knowing which toxicology tests are best are predicting human toxicity and, if so, how much. Fundamental changes in toxicology are required to get data for benefits estimation. In my judgment, the pressure for estimating the benefits and costs of abating hazardous air pollution is putting the right sort of pressure on the regulatory environment. These pressures will force changes in toxicology that will improve identification and quantification of human toxicity.

## Introduction

My first task is to inform risk analysts what economists need to estimate the benefits and costs of abating hazardous air pollutants. That is a straightforward task. My second task is to tell economists why the current toxicology information that they are getting is not capable of estimating these benefits. My third task is to open a dialogue between risk assessors and benefits estimators that might some day result in confident estimates of the benefits. My final task is to suggest a research agenda leading to confident estimates of the benefits of abating hazardous air pollutions. To illustrate the first task, I begin with a review of a large EPA benefit-cost analysis.

## Assessing the Benefits and Costs of the Clean Air Act

Assessing the benefits and costs of an environmental program, such as the retrospective and prospective effects of the 1970 Clean Air Act, requires information beyond what is factually available. Estimating the costs of the act requires estimating the private and social costs, including abatement costs and opportunity costs, of the regulations. This cost must be contrasted with the costs that would have been incurred if the Clean Air Act and associated regulations did not exist. The actual costs incurred are difficult to estimate; the costs that would have been incurred absent the act cannot be observed or even estimated with confidence, since the world of no regulation does not exist. The cost estimation is, by far, the easier part of the benefit-cost analysis.

The first step in assessing the benefits is to estimate the improvement in air quality

resulting from the Clean Air Act. Unfortunately, there are not good inventories of the ambient air quality or the amounts of each pollutant emitted even today, much less the ambient air quality and amount emitted in 1969. Since the economy has grown since 1970, we need to estimate ambient air quality and how much emissions would occur at each date, with and without the Clean Air Act. This task is similar to that for estimating the costs of abatement.

The second step in estimating benefits is estimating the social benefits that result from lower air pollution levels. The benefits include human morbidity and mortality, visibility, odors, aesthetics more generally, damage to ornamental plants and crops, and other effects from controlling air pollutants. In particular, the following steps are necessary:

Identify each relevant category of harm – eliminate those that are "trivially" small.
Quantify the relationship between ambient air quality and each effect.
For each year, assess the changes in ambient air quality as a result of the regulations.
For each year, estimate the physical benefits of cleaner air in terms of the categories identified in 1 and quantified in 2, e.g., premature deaths, cases of each disease, quality adjusted life-years or disability-adjusted life-years, better visibility, etc.
Value these estimated physical benefits in dollars or compute the value of each relative to the others and compute a benefits index.
Compute the net benefit (benefits minus costs) if the benefits are valued in dollars or the cost-effectiveness (dollar costs divided by the benefits index) if only a benefits index can be computed.

Many non-economists think that 5 is the hardest step, the one that cannot be done rigorously. That is wrong. Methods have been developed to estimate these values. Although there are important uncertainties associated with each method, valuation is not the main source of uncertainty in estimating benefits.

Step 3 adds more uncertainty than step 5, in general. We don't know how the exposure of the average American has changed as a result of the CAA. We have measurements of the criteria air pollutants in a few hundred locations, but translating this information into what people breath when they are outside is difficult. Still more difficult is estimating the amount they breathe during the day, since they spend only a small amount of time outside of buildings or motor vehicles. Additional problems are the health interactions of pollutants, sensitized individuals, and local high pollution concentrations.

Estimating Effects on Humans: Using Epidemiology and Toxicology Information

But the greatest uncertainty is contributed by Step 2. Epidemiology studies the morbidity and mortality in humans that result from exposure to toxicants. Unfortunately, there are a host of problems with the usual epidemiology study. The data are almost always observational rather than experimental. Epidemiologists are tempted to search for associations that were not hypothesized before the study began. The data sets always have important uncontrolled factors influencing the results. The data are imprecise or incomplete because of people lost to follow-up or misdiagnosis. The dose (exposure) is

almost never known with prevision, there are interactions with other environmental exposures and genetic defects, and the full extent of reactions among people who are exposed is unknown. Subtle reactions are almost impossible to identify. For example, some epidemiologists use a relative risk of two as the criteria for having confidence in an observed relationship.

Connoisseurs of epidemiology bemoan the misleading studies and quantification difficulties. More than one has told me that they would prefer to rely on some other approach. That statement reminds me of the story of someone seeking to hire a personal assistant. With two persons to choose between, he interviewed the first and immediately hired the second.

Despite these formidable difficulties, epidemiology observes morbidity and mortality in humans. Toxicology generally studies the effects of a toxicant on a laboratory animal or a cell culture. Occasionally, human volunteers to are exposed to the toxicant at levels that are believed too low to harm the individual. For ethical reasons, the studies with human volunteers keep the dose below the level that would be expected to cause even a small adverse effect. Thus, there is no information on disease; one must extrapolate from an observed effect (not adverse) from an acute dose to disease at a higher dose that is usually present for a long period of time. Despite these formidable difficulties, studies on humans obviate the need to extrapolate from rodents or from cell cultures. These studies can provide precise measurement of the effects of the toxicant in these setting, but say nothing directly about what will be the effect on human morbidity and mortality from relevant exposures.

Problems such as lack of standardization of protocols are small compared to the central problem of toxicology: In only a few cases have the results of in vivo or in vitro studies been compared with human outcomes for the test substance. It is hard to know how to interpret the results of a toxicology study. A well-done study will have high "internal" validity, meaning that it can be replicated within this lab and even in other labs. However, the study has little or no "external" validity in the sense of knowing its implications for human health.

Regulatory toxicology seeks to protect humans against harmful exposure to toxicants. Since it is difficult to draw inferences about the harm to humans from experiments with rodents or cell cultures or to extrapolate from high doses to low doses, without knowing the physiological mechanisms by which exposure to a toxicant causes disease, toxicologists have made a set of "conservative" assumptions that are designed to give a "plausible upper bound" to human toxicity. For example, for cancer the standard assumption is to construct the 95% upper confidence level for the exposed rodents with the steepest dose-response relationship. Other assumptions are made about the shape of the dose-response relationship and the exposure of the population that intended to make sure that the risk to humans is not underestimated.

Hormesis

Another major issue is the large body of data showing that low level exposures to toxicants may improve, rather than harm, health. The usual assumption for cancer risk assessment is that even a single molecule of a carcinogen has a small chance of causing cancer. A great deal of laboratory data on animals and even some data on humans suggests that a tiny dose might improve health.

Hormesis becomes a dominant issue for policy in cases, such as benzene, where most Americans are exposed to benzene at parts per billion concentrations. Most of the estimated cases of leukemia from benzene exposure occur at a few parts per billion. Hormesis suggests that concentrations at this level might improve health, rather than cause leukemia.

Cancer Risk Assessment

Figure 1 illustrates a problem in interpreting toxicology data. These data summarize the outcome of National Toxicology Program lifetime rodent cancer bioassay results for about 1,000 chemicals. A first way of looking at the data is the concordance between rats and mice: 70%. Thus, while there is general agreement between rats and mice, the agreement is far from perfect. If we flipped a coin in order to predict the rat outcomes, we would have a concordance of 50%.

Figure 1

| | | Rats | |
|---|---|---|---|
| | Carcinogen? | Yes | No |
| Mice: Carcinogen? | | | |
| | Yes | 35 (TP) | 15 (FP) |
| | No | 15 (FN) | 35 (TN) |

Another view of the data is the ability of a mouse test to predict the carcinogenicity of each chemical for rats. Viewed that way, when the bioassay is positive on mice and is also positive for rats, that is a "true positive" or TP. Unfortunately, for some chemicals the positive result for mice is negative for rats. This "error" is a "false positive" or FP. When the bioassay is negative in mice and negative in rats, that is a "true negative" or TN. Unfortunately, some chemicals are negative in mice but positive in rats, a "false negative" or FN. The figure shows that, for chemicals that are positive in mice, 70% of them are positive in rats (TP) and 30% are negative in rats (FP). Of chemicals that are negative in mice, 70% are negative in rats (TN) and 30% are positive in rats (FN). The usual interpretation of the NTP is that a chemical that is positive in either rodent species is considered to be a possible human carcinogen. Thus, 65% of chemicals tested are considered to be possible human carcinogens.

These NTP bioassay results are used to classify chemicals as likely human carcinogens. Given the 70% concordance between rats and mice, what is the likely concordance between rodents and humans? Since rats are more similar to mice than either are to

humans, it seems likely that the concordance between rodents and humans will be less than 70%. What proportion of the 65% of chemicals that are positive in rodent tests are human carcinogens? How many false positives and false negatives are there likely to be in this classification? As noted above, we know that there are some false positives (the alpha-2-u globulin chemicals) and some false negatives (benzene). The NTP interpretation is intended to minimize the number of false negatives, even at the cost of additional false positives.

It seems unlikely that the concordance between rodents and humans is as high as 70%. In fact, it is more than just possible that the concordance between rodents and humans is no better than 50%, equivalent to a coin flip, which is a bit cheaper than spending more than $1 million on a lifetime rodent bioassay.

This is relevant because of the limited data on toxicity in humans. For example, there is epidemiology data connecting each chemical to a cancer for about two dozen chemicals or groups of chemicals. Unfortunately, the concentrations that people were exposed to are often highly uncertain and so there is only a remote idea of the dose-response relationship. For the other 600 plus toxicants in the Toxic Release Inventory report, there is at best toxicology data, generally in the form of rodent studies. Biologists warn that extrapolating between species is perilous. Using toxicological data to estimate the risks to humans from exposure to a toxicant does precisely that extrapolation. The extrapolation begins with a leap of faith that the effect observed in rodents or in cultured cells predicts human toxicity. The differences in anatomy and physiology between humans and rodents means that many diseases/conditions are unique to rodents or to humans. Still more uncertain is extrapolating human risk from the dose-response relationship observed in rodents.

Non-cancer Risk Analysis

Non-cancer risk analysis proceeds by establishing the no observable effects level (NOEL) in rodents and extrapolating to humans by using safety factors to account for differences among species and for the most sensitive individuals. The practice assumes that humans are the most sensitive species, despite considerable data showing that other species are often more sensitive, e.g., dioxin in mice.

Several proposals have been published to estimate harm for exposure above the reference dose or to make use of a combination of data and judgment. However, in none of this work is there an attempt to give an unbiased estimate of the effect on humans. Past data suggest that for both cancer and other health effects rodents sometimes are positive when humans are negative and vice versa. Furthermore, when both are positive, sometimes the rodents are more sensitive than humans and vice versa. These data suggest that if one knew no more than the results of an in vivo experiment, one would be very uncertain about the implications for humans, both in terms of whether the chemical is a human toxicant and, if so, what is the potency.

I don't mean to suggest that benefits estimators are prissy, unused to highly uncertain data. Indeed, anyone who has read the benefits assessment from the 812 report might be

inclined to ask, as an epidemiologist did of me some years ago: "Are there any data so bad that an economist would not analyze them?" However, the economist needs unbiased estimates, even if they are highly uncertain. Toxicologists are not providing unbiased estimates - that is the fatal flaw.

A Research Agenda

Risk analysis could be most helpful to benefits assessment if toxicologists performed the following analysis:

For cancer, for all the known human carcinogens, examine whether a standard National Toxicology Program bioassay would be positive. If there is some uncertainty, what is the likelihood that a particular bioassay would be positive?

Compare the estimated dose-response relationship for rodents with the relationship for humans, accounting for the uncertainty in the resulting estimates.

1. For chemicals that are positive in the NTP bioassay, which are not or are unlikely to be human carcinogens?
2. For non-cancer endpoints, the same questions are relevant, although there is no single standard for in vivo or in vitro tests. In particular, what is the concordance between different species in the same test and across different in vivo and in vitro tests? What is the concordance between test outcomes and human toxicity data on each chemical?

The other issue is extrapolating from high human doses, observed in accidents, occupational exposure, or tests with human volunteers, as well as extrapolating from high doses in in vivo and in vitro tests to the low doses over long periods that most people experience. In the absence of knowing the mechanism of action, one must rely on assumptions about the nature of the dose-response relationship. The ED01 experiment attempted to pin down the best dose-response relationship for cancer. Unfortunately, all of plausible models performed about as well in explaining the observed data, even though they had very difference implications for the effects at low exposures.

Summary and Conclusions

As currently practiced, regulatory toxicology cannot provide data to estimate the benefits of abating hazardous air pollutants. No minor patches will provide these data. Regulatory toxicology is built on a foundation of protecting humans from harmful exposure to toxicants. It is saturated with implicit and explicit assumptions that lead to "plausible upper bounds" rather than "best estimates" of the exposure-response relationship.

If we knew the mechanisms of action by which toxicants harm humans, we could establish standards that would protect them and also provide best estimates of the exposure-response relationships to the benefits assessors. At present, few mechanisms of action are known and it seems doubtful that we will ever know the mechanisms of action for most toxicants.

Short of knowing the mechanisms, research can do much to clarify the qualitative and

quantitative risks to humans of exposure. Toxicologists need to analyze the implications of current in vivo and in vitro tests for human toxicity. They need to look for human data to compare with laboratory results. I have no doubt that when this happens, we will find that some popular tests predict human toxicity no better than flipping a coin. For these tests, society is wasting its resources and using meaningless data to make regulatory decisions. Other tests can be modified to increase their human predictivity. New tests can be developed that are more predictive of humans.

Insisting on estimating the benefits of reducing exposure to hazardous air pollutants will lead to better policy. More important, it should trigger a revolution in toxicology in searching for laboratory tests that are more predictive of human toxicity.